

Development of Voice Control Algorithm for Robotic Wheelchair Using NIN and LSTM Models

Mohsen Bakouri^{1,2,*}

¹Department of Medical Equipment Technology, College of Applied Medical Science, Majmaah University, Majmaah City, 11952, Saudi Arabia

²Department of Physics, College of Arts, Fezzan University, Traghan, 71340, Libya

*Corresponding Author: Mohsen Bakouri. Email: m.bakouri@mu.edu.sa

Received: 12 November 2021; Accepted: 09 February 2022

Abstract: In this work, we developed and implemented a voice control algorithm to steer smart robotic wheelchairs (SRW) using the neural network technique. This technique used a network in network (NIN) and long short-term memory (LSTM) structure integrated with a built-in voice recognition algorithm. An Android Smartphone application was designed and configured with the proposed method. A Wi-Fi hotspot was used to connect the software and hardware components of the system in an offline mode. To operate and guide SRW, the design technique proposed employing five voice commands (yes, no, left, right, no, and stop) via the Raspberry Pi and DC motors. Ten native Arabic speakers trained and validated an English speech corpus to determine the method's overall effectiveness. The design method of SRW was evaluated in both indoor and outdoor environments in order to determine its time response and performance. The results showed that the accuracy rate for the system reached 98.2% for the five-voice commands in classifying voices accurately. Another interesting finding from the real-time test was that the root-mean-square deviation (RMSD) for indoor/outdoor maneuvering nodes was 2.2×10^{-5} (for latitude), while that for longitude coordinates was a whopping 2.4×10^{-5} (for latitude).

Keywords: Network in network; long short-term memory; voice recognition; wheelchair

1 Introduction

Disabled people in public places have a complex time maneuvering wheelchairs. Those people also depend on others to assist them in moving their wheelchairs [1]. According to [2], People with limited mobility make up 40% of those unable to steer and maneuver wheelchairs adequately, compared to the 9%–10% who have been taught to operate power wheelchairs. Furthermore, clinical studies indicate that nearly half of 40% of disabled with impaired mobility cannot control an electric wheelchair. More than 10% of those disabled who use electric wheelchairs have had an accident within the first four months [3]. Accordingly, to provide a better quality of life for wheelchair users, it has been developed



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

with various technologies fitted with a navigation and sensor system that works automatically [4–7]. These wheelchairs are known as Smart Robotic Wheelchairs (SRW) due to the introduction of more choices for controlling the chair, improved safety, and comfort over conventional wheelchairs [8,9].

Generally, to perform autonomous activities, the SRW must be capable of navigating safely, avoiding obstacles, and passing through doorways or any other confined space [10]. In SRW, which is controlled via a joystick intelligent control system unit, the operation of this system has proven to be the most significant development [11]. However, people with disabilities in their upper extremities will have difficulty using the joystick smoothly. As a result, situations requiring quick response could result in tragic incidents [12]. Therefore, several researchers have started to develop SRW based on human physiological signals. For example, the human-computer interface (HCI) operates a wheelchair by the use of physiological signals such as the electrooculogram (EOG), the electromyogram (EMG), and the electroencephalogram (EEG) [13–15]. On the other hand, brain-computer interfaces (BCIs) have advantages for translating brain signals into action for wheelchair control [16]. The hybrid BCI (hBCI) approach, which integrates EEG and EOG, increased wheelchair accuracy and speed. However, the technology of EEG-BCI has several limitations in terms of low resolution and signal-to-noise ratio (SNR). Additionally, hBCI encounters difficulties simultaneously controlling speed and direction [17–19].

In general, different researchers have significantly enhanced the development of SRWs with autonomous functions via voice recognition technology. The strategy described in [20] illustrated the result of an intelligent wheelchair system using a voice recognition technique in conjunction with a GPS tracking model. By using a Wi-Fi module, voice commands were transformed into hexadecimal number data and used to drive the wheelchair in three different speed phases. Additionally, the system utilized an infrared (IR) sensor to identify barriers and a mobile application to determine the patient's location. A similar work conducted by Raiyan et al. [21] utilizes an Arduino and Easy VR3 with a voice recognition module to drive an autonomous wheelchair system. This study demonstrated that the implemented system robustly guided the wheelchair with less complex data processing and without wearable sensors. A different novel study employs an adaptive neuro-fuzzy to steer a motorized wheelchair [22]. The study was created and executed using real-time control signals supplied by voice instructions via a classification unit. This architecture's proposed system for tracking the wheelchair uses a wireless sensor network [22]. Despite the highly advanced methodologies presented by researchers in this field, the high cost, and precision required for distinguishing, categorizing, and identifying the patient's voice continue to be the primary obstacles.

Recently, numerous researchers have employed the convolutional neural network (CNN) technology to overcome the inaccuracy of identifying and classifying patients' speech [23,24]. This technology converts speech commands into spectrogram visuals, then fed into CNN. This technique has been shown to improve the accuracy of speech recognition. In this context, Sharifuddin et al. [25] introduced an inelegant design using CNN to steer SRW based on four voice commands. The method used data collected from the google website and applied Mel-frequency cepstral coefficient (MFCC) to extract voice commands. Authors claim that the results of the vice commend classification using CNN have an accuracy of 95.30%. Similarly, Sutikno et al. [23] developed a voice-controlled wheelchair-using CNN and long short-term memory (LSTM) based on five commends. The developed method used data obtained from recording several subjects using sound recorder pro and sox sound exchange. This method demonstrated that the vice commands classification using CNN and LSTM has accuracy above 97.80%. Although many of the research results that have been conducted are significantly high, computers are still used in these methods to perform complex operations on CNN.

However, due to the extensive calculations required to attain high accuracy, employing CNN in smartphones is still developing [26]. This article proposes to design a voice control method for robotic wheelchairs by using CNNs and LSTM models [27]. The system used a smartphone to build an interactive user interface that can be controlled easily by delivering a voice command to the system’s motherboard via the mobile application. The objective of this study was accomplished by developing and implementing a mobile application, a voice recognition model, a CNN model, and an LSTM model. Additionally, all safety considerations were taken into account while driving and navigating in both indoor and outdoor environments. The results indicated that the built system was remarkably resilient in terms of response time and correct execution of all orders without delay.

2 Materials and Methods

2.1 Architecture of the System

The implementation of the proposed architecture system is divided into two stages, as illustrated in Fig. 1. In the first stage, an Android mobile application was developed using the Flutter programming language [28,29]. Six steps were used to create and program the application (named voice control), as shown in Figs. 2a and 2b. The voice control application appears in the application list when accessed. After granting the application access to the microphone, it attempts to recognize the words and highlights them in the interface recognition, as depicted in Fig. 2c. The second stage consists of assembling hardware devices, including mechanical parts and a control unit, as shown in Fig. 3. The mechanical parts are composed of a standard wheelchair, two motor pairs (3.13.6 LST10 24v DC 120 rpm), and an NP7–12 12v 7ah lead acid battery. At the same time, the control unit includes Raspberry pi4 (GPU: Broadcom Video Core VI, Networking: 2.4 GHz, RAM: 4 GB LPDDR4 SDRAM, Bluetooth 5.0, microSD), and Relay Module (5 V 4-channel relay interface board).

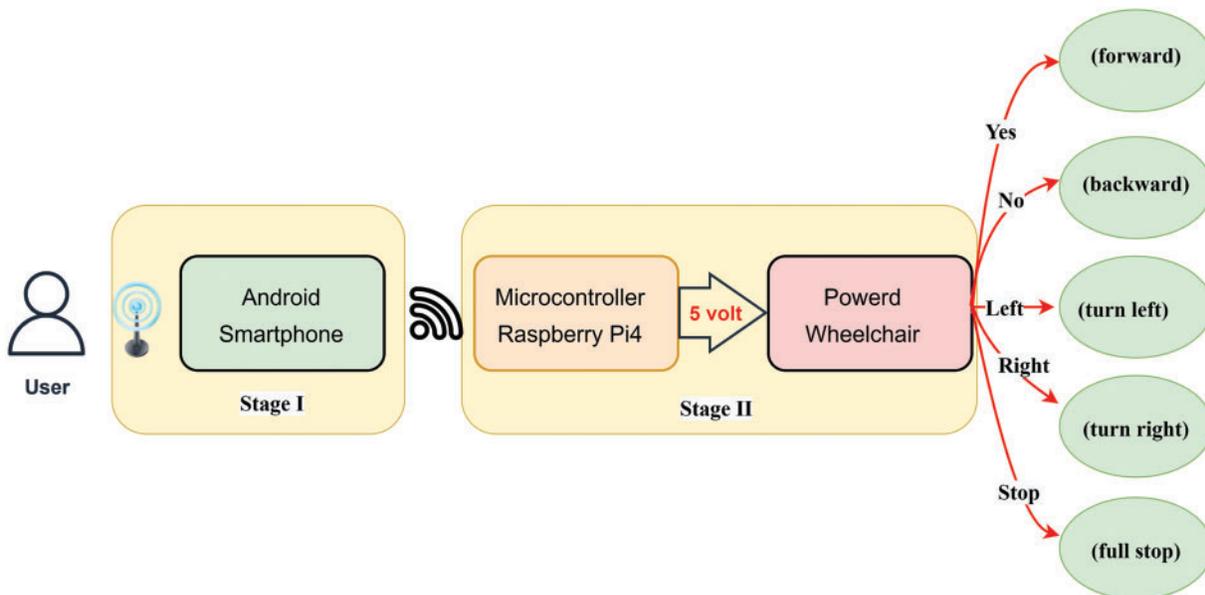


Figure 1: Overall system architecture

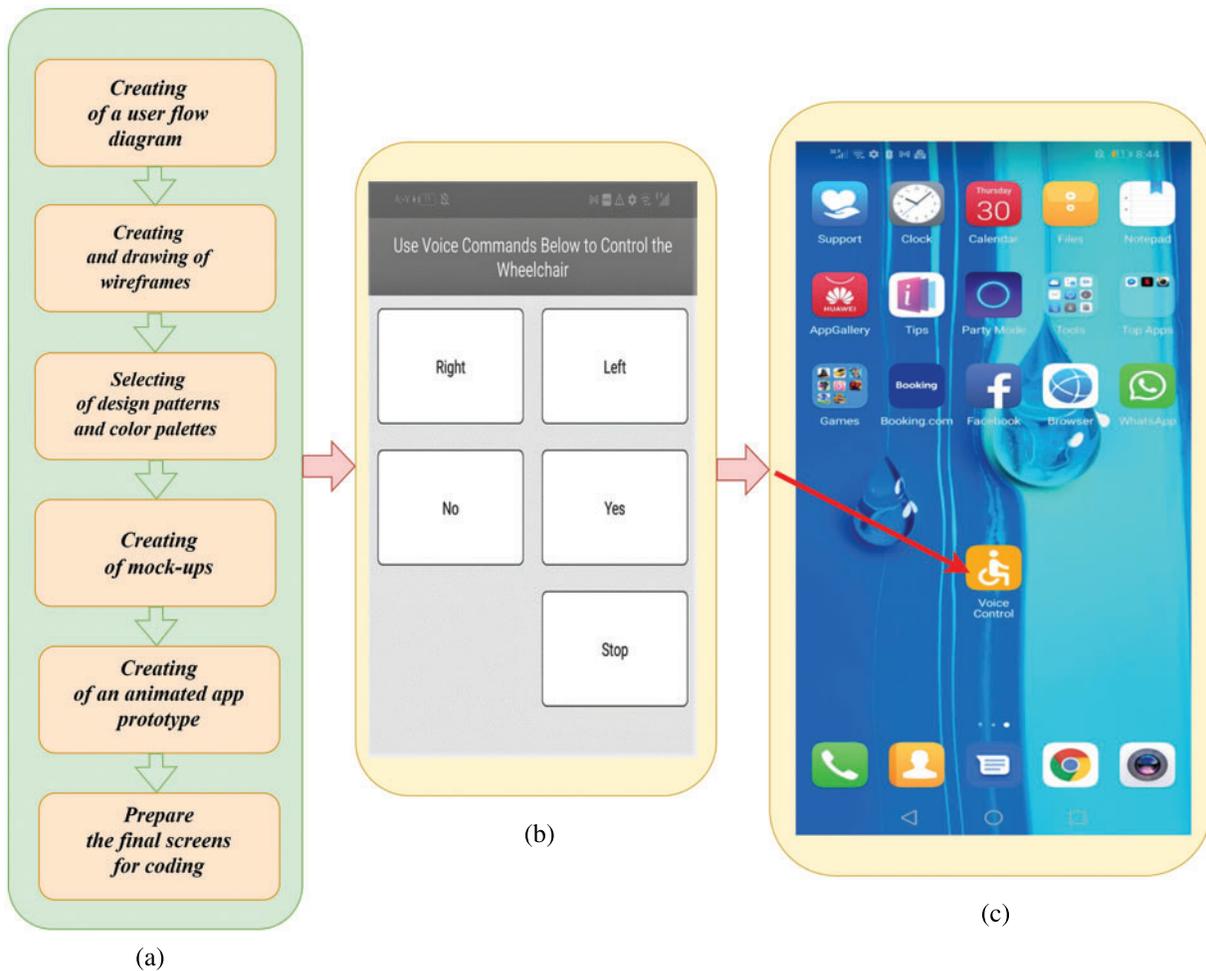


Figure 2: Android app interface created for controlling powered wheelchair and its voice command prediction ratio: (a) Steps of creating the application (a) Main of voice control app, (b) Mode screen (Right, Left, No, Yes, and stop)

2.2 Development of Voice Recognition Model

Feature extraction is used to produce a frequency map for each audio file, which displays how the signal evolves over time. As a result, speech analysis systems used the Mel-frequency cepstral coefficients (MFCC) coefficients to extract this information [30]. An important part of character extraction is preventing numerical instability by putting it through a finite impulse response filter (FIR), which is a single-coefficient, digital filter as:

$$q(n) = v(n) - \xi v(n - 1) \quad (1)$$

where $q(n)$ and $v(n)$ are the output filter and the original voice signal respectively. The number of sampling is denoted by n , and ξ given as $0 < \xi \leq 1$.

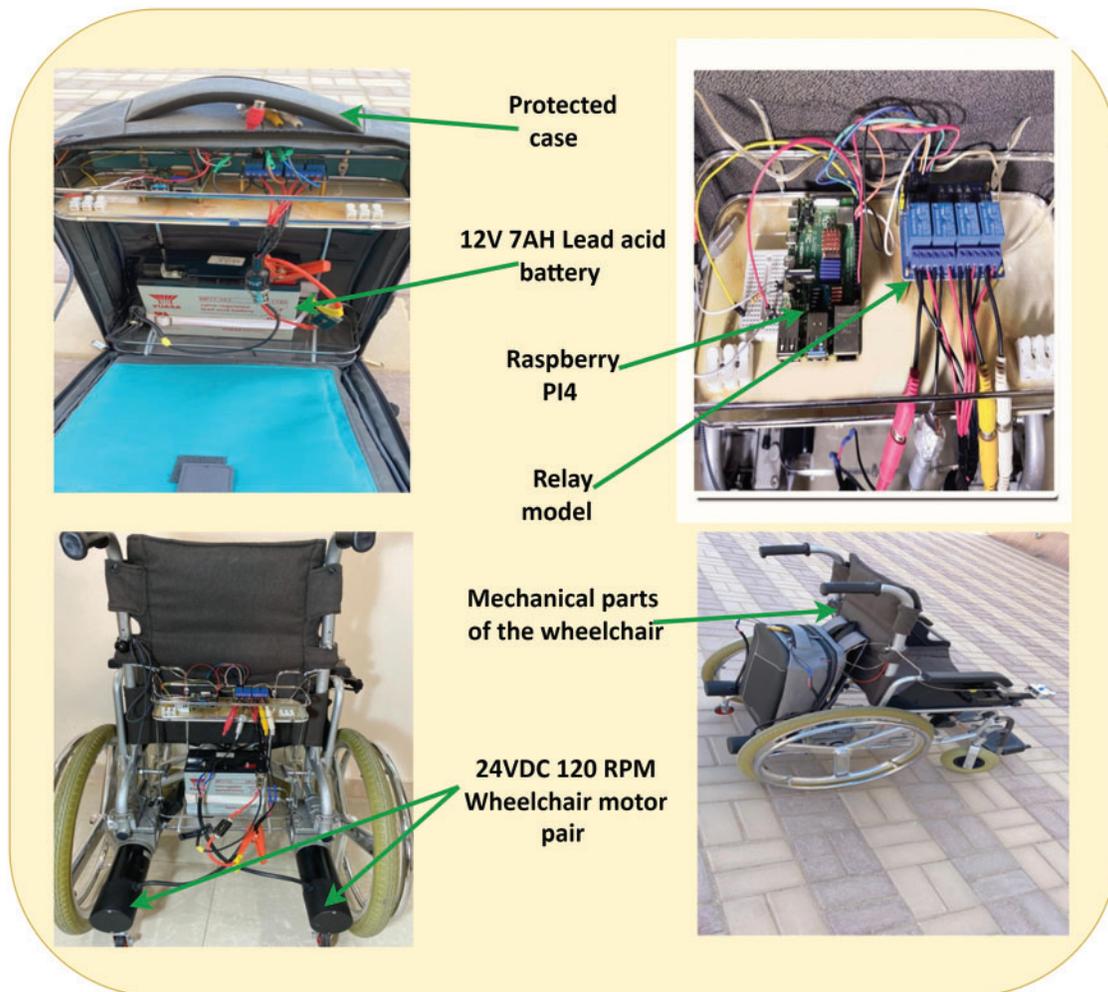


Figure 3: Mechanical assembly of wheelchair

It is necessary to use framing and windowing $[w(n)]$ in order to maintain the samples within frames and reduce signal discontinuities as:

$$w(n) = \begin{cases} (1 - \alpha) - \alpha \cos\left(\frac{2\pi n}{N-1}\right) & n = 0, 1, \dots, N-1 \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

where α is a constant, and N represent the number of frames.

In this method, the spectral analysis is achieved using fast Fourier transform (FFT) to calculate the magnitude spectrum for each frame as:

$$q(k) = \sum_{n=0}^{N-1} q(n)e^{-\frac{j2\pi kn}{N}}, \quad k = 0, \dots, N-1 \quad (3)$$

The spectrum is subsequently processed according to MFCC using a bank of filters; where Mel-filter-bank can be written as:

$$H_m[f_k] = \frac{f_k - f[m-1]}{f[m] - f[m-1]} \quad (4)$$

It is possible to write the boundary points ($f[m]$) by taking the lowest (f_l), and highest (f_h) values on the filter-bank in terms of hertz and frequency as:

$$f[m] = \left(\frac{N}{F_s}\right) B^{-1} \left(B(f_l) + m \frac{B(f_h) - B(f_l)}{M+1} \right) \quad (5)$$

where M , and N are the number of filters and the size of the FFT respectively. The term B is representing the Mel-scale which calculated by:

$$B(f) = 1125 \ln\left(1 + \frac{f}{700}\right) \quad (6)$$

In this work, we used an approximate homomorphic transform to remove noise and spectral estimation errors as:

$$S[m] = \ln \left[\sum_{k=0}^{N-1} |q[k]^2 H_m[k]| \right], \quad 0 < m \leq M \quad (7)$$

In the final step of MFCC processing, Cosine Transformer (DCT) are employed to provide high decorrelation properties as:

$$c_l[n] = \sqrt{\frac{2}{M}} \sum_{m=1}^M S_l[m] \cos \left[\frac{n\pi}{M} \left(m - \frac{1}{2} \right) \right], \quad n = 0, 1, \dots, L < M \quad (8)$$

The first and second derivatives of (8) are used to obtain the feature map:

$$\Delta c_l[n] = \frac{\sum_{p=1}^p P (c_{l+p}[n] - c_{l-p}[n])}{2 \sum_{p=1}^p P^2}, \quad \Delta^2 c_l[n] = \frac{\sum_{p=1}^p P (\Delta c_{l+p}[n] - \Delta c_{l-p}[n])}{2 \sum_{p=1}^p P^2} \quad (9)$$

The database was thus developed and used by CNN, and it applies to all recordings that have been recorded.

2.3 Development of CNN Model

This study used the NIN structure as the core architecture for developing mobile applications [27,31]. NIN is a CNN technique that does not employ fully connected (FC) layers. Instead, NIN uses global pooling rather than fixed-size pools to take images of any size as inputs. This technique is helpful for mobile applications because it allows users to fine-tune the speed-accuracy trade-off without compromising network weights.

In order to develop CNN, we employ a multi-threading technique. The smartphone used in this technique has four CPU cores, making it simple to divide a kernel matrix into four sub-matrices and divide a row into four sub-matrices. To obtain the output feature maps, it is necessary to conduct four

generalized matrix multiplication (GEMM) operations simultaneously. The cascaded cross channel parametric pooling (CCCPP) technique was also utilized to compensate for the loss of the FC layers. Because of this, our CNN model comprises input and output, twelve convolution layers, and two succeeding layers.

2.4 Development of LSTM Model

We adopted LSTM model as a vanilla structure [32]. The architecture of this model consists of a set of recurrently connected sub-networks, known as memory blocks. In specific, the model is composed of a cell, an input gate, an output gate, and a forget gate. The mechanism of the LSTM model start works by identifying and eliminating of last outputs data (h_{t-1}) and current inputs data (X_t) using the sigmoid function (σ). This step is achieved by forgetting gate (f_t) function, which is given by:

$$f_t = \sigma(W_f[h_{t-1}, X_t] + b_f) \quad (10)$$

where W_f and b_f are weight matrices and bias weight vector respectively. In the second step, the model will store and update the new input data in the cell state (C_t) using the sigmoid layer and tanh layer. The sigmoid layer decides to update or neglect the new data using (1 or 0), while the tanh layer gives weighs to the passing data using (1 or -1). Then, the old memory data (C_{t-1}) added to the new memory of the cell state as:

$$\begin{aligned} i_t &= \sigma(W_i[h_{t-1}, X_t] + b_i) \\ K_t &= \tanh(W_c[h_{t-1}, X_t] + b_c) \\ C_t &= C_{t-1}f_t + K_t i_t \end{aligned} \quad (11)$$

In the final step, the output value (h_t) is calculated based on the sigmoid gate (O_t) and the new values created by C_t and tanh layer as:

$$O_t = \sigma(W_o[h_{t-1}, X_t] + b_o) \quad h_t = O_t \tanh(C_t) \quad (12)$$

Fig. 4 illustrates the diagram of the proposed CNN with LSTM neural network.

The flowchart in Fig. 5 depicts the signal flow for controlling wheelchair system.

3 Experimental Procedure

In the Health and Basic Sciences Research Center at Majmaah University, the English speech corpus of isolated words was used to evaluate the proposed system. Ten native Arabic speakers were selected to pronounce five words with a total of 2,000 utterances. The data was recorded using a 20 kHz sample rate and 16-bit resolution. Then, using the reinforcement method, this data set was supplemented with additional audio cues. The supplementary dataset contains 2,000 speech altered in pitch, velocity, dynamic range, noise, and forward and backward time shift. The new data set (original and supplemented) contains 4000 utterances and is divided into two parts: a training set (training and validation) containing about 80% of the samples (3200), and a test set containing the remaining 20% of the sample (800).

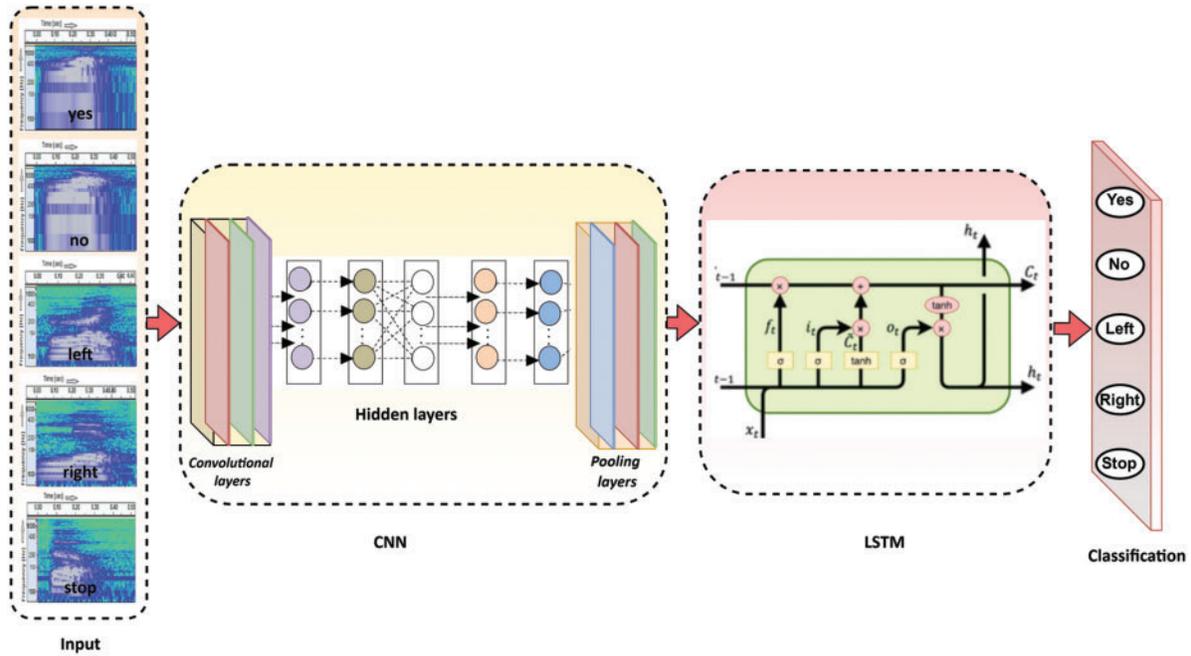


Figure 4: The CNN with LSTM neural network

To quantify the predictive accuracy and quality for the proposed system, we compute the F-score as:

$$F - score = 2 \left(\frac{P * R}{P + R} \right) \quad (13)$$

where P and R indicate for precision and recall, respectively, and are defined as follows::

$$P = \frac{T_p}{T_p + F_p} \quad (14)$$

$$R = \frac{T_p}{T_p + F_N} \quad (15)$$

here, F_p , F_N , and T_p are false positive, false negative, and true positive respectively.

During the classification, the percentage difference (%d) equation was employed to measure the accuracy of each voice command prediction as:

$$\%d = \frac{|V_1 - V_2|}{\left(\frac{V_1 + V_2}{2} \right)} * 100\% \quad (16)$$

where V_1 and V_2 represent the first and second observations during the comparison process. Indoor/outdoor navigational performance is also assessed in real time using this methodology. With vocal commands, the wheelchair navigated around and inside a mosque at 24.893374, 46.614728 coordinates.

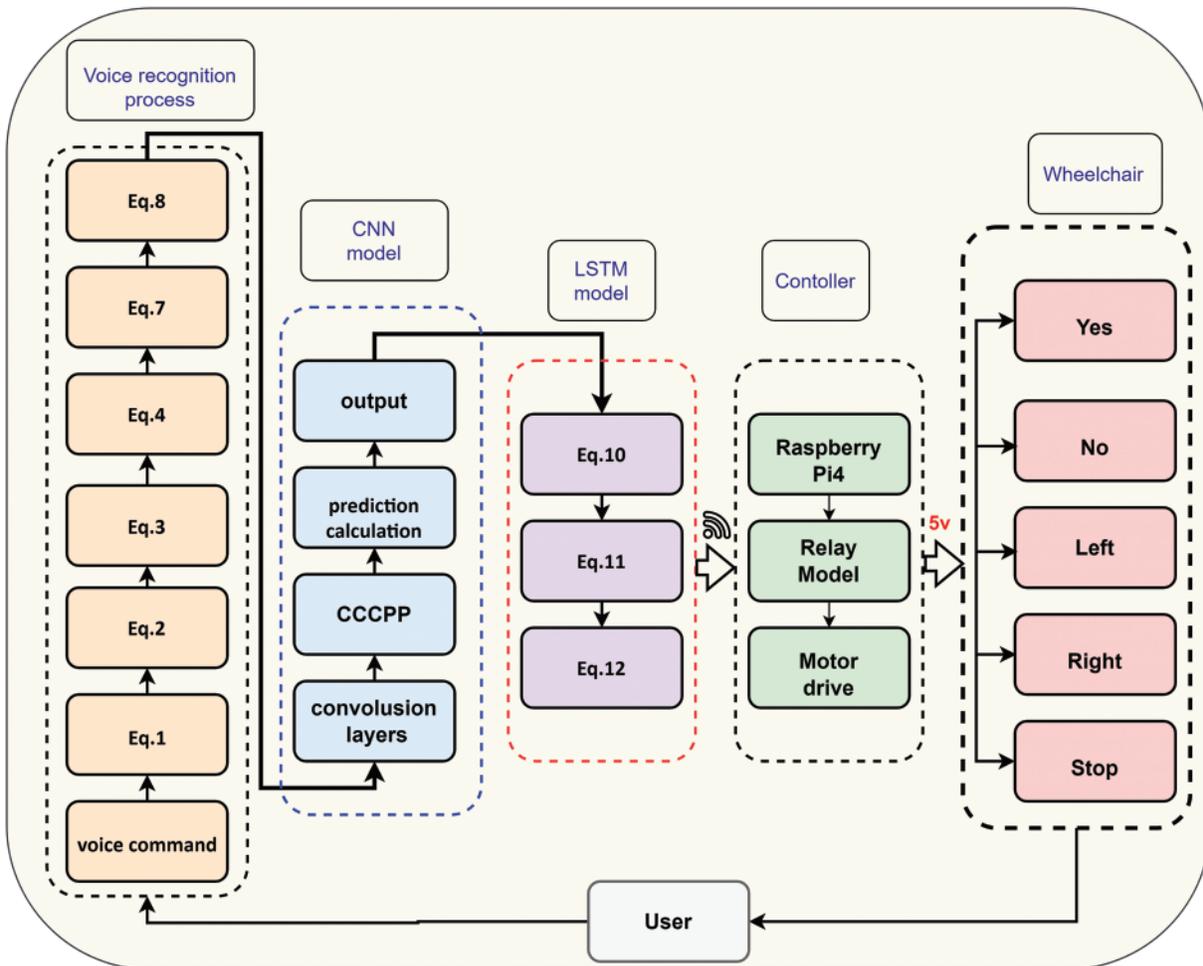


Figure 5: Flowchart of the system

4 Results

Tabs. 1–3 represent the steps of voice recognition model development. Tab. 1 shows the five audio wave shapes with training time.

The next step is to convert the audio file waves into its frequency domain by using Fourier analysis as shown in Tab. 2.

Then, these frequency domain waves were converted into spectrograms (See Tab. 3) and used as input in NIN model then to LSTM model respectively.

Table 1: The five audio wave shapes

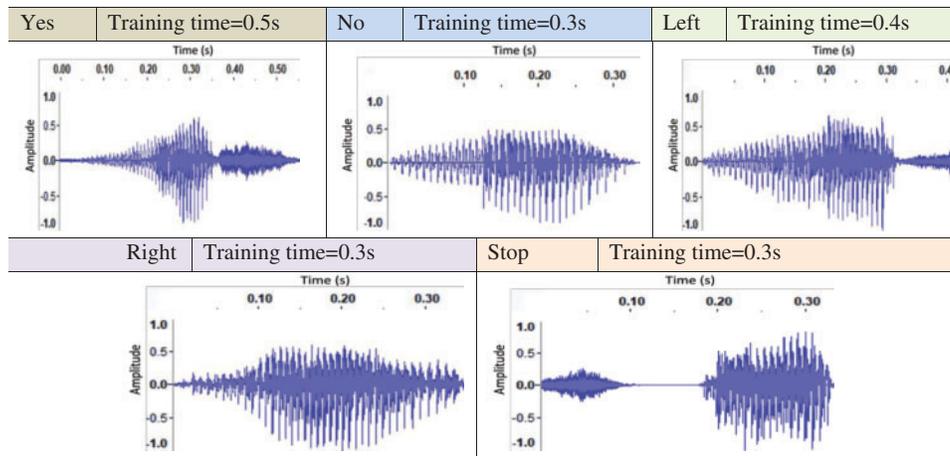


Table 2: The five singles of long-term spectrum

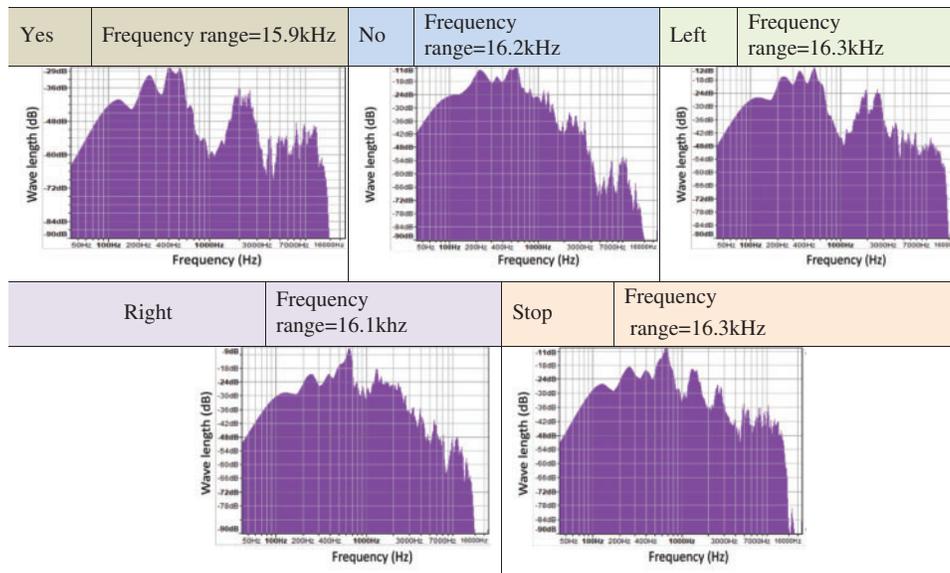
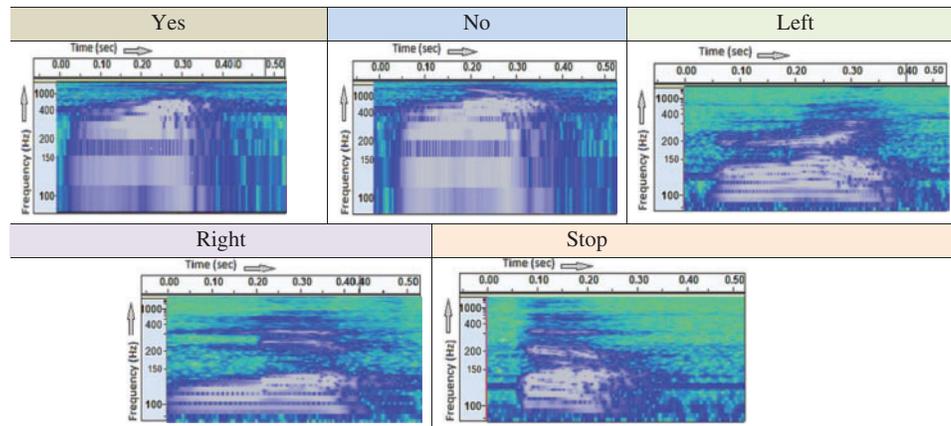
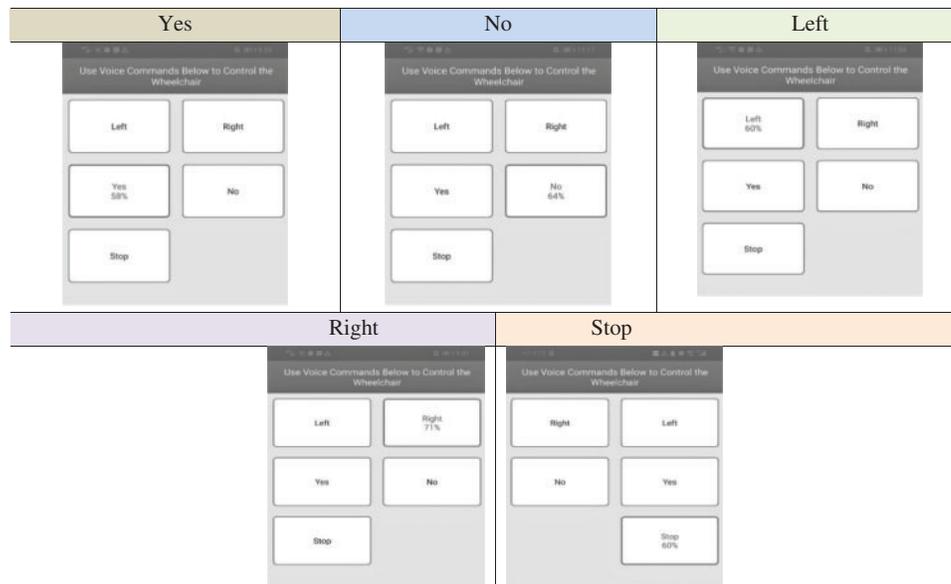


Table 3: The five singles of time-frequency spectrogram



Tab. 4 illustrated the screen shoot of mobile application (voice command prediction ratio). Additionally, the application displays the user’s expected word weight. It is usually a single-voice order with greater weight than other words, indicating that no incorrect classification judgment can be made during the classification process.

Table 4: The five words of screen shoot for mobile app



The confusion matrix was generated using the initial results, as shown in [Tab. 5](#). The average accuracy was around 82.6% of the accurate forecast for five-voice commands. We used the phrases true positives, true negatives, false positives, and false negatives to describe the classification activity. The computations for the voice-command prediction ratio, accuracy, and precision are shown in [Tabs. 6](#) and [7](#). In terms of calculating the percentage difference between two commands when comparing them, the example compares “STOP” to other commands. This suggests a minor risk of selecting an inaccurate classification option. On the other hand, the difference between accurate and erroneous predictions is quite significant, indicating a negligible risk of making incorrect predictions. The difference exceeded 187 percent, as shown in [Tab. 7](#).

Table 5: Normalized of confusion matrix

		Actual voice command				
Prediction ratio %	Class	Yes	No	Left	Right	Stop
	Yes	87%	3%	3%	4%	3%
	No	2%	90%	3%	2%	3%
	Left	4%	1%	92%	2%	1%
	Right	5%	4%	2%	87%	2%
	Stop	3%	5%	3%	3%	86%

Table 6: Accuracy, precision, recall, and F-score for voice commands

Class	Accuracy	Precision	Recall	F-score
Yes	95%	0.73	0.74	0.735
No	96.3%	0.75	0.75	0.75
Left	98.2%	0.77	0.75	0.76
Right	94.8%	0.75	0.73	0.74
Stop	93.5%	0.71	0.69	0.70

Table 7: Calculation of percentage difference for stop command

Voice command	Yes	No	Left	Right	Stop
Prediction ratio	3%	5%	3%	3%	86%
Percentage difference	187%	178%	187%	187%	—

An evaluation of indoor/outdoor navigation for the wheelchair was obtained to test the real-time performance of the designed system within the public area. [Fig. 6](#) depicts the intended route navigation in comparison to the actual path. [Tab. 8](#) shows the coordinate nodes of the intended and actual pathways while traversing. The root means square deviation (RMSD) was used to represent the difference between the planned and actual nodes in this experiment. Figures show that the RMSD for latitude and longitude coordinates are $2.2 * 10^{-5}$ and $2.4 * 10^{-5}$, respectively.

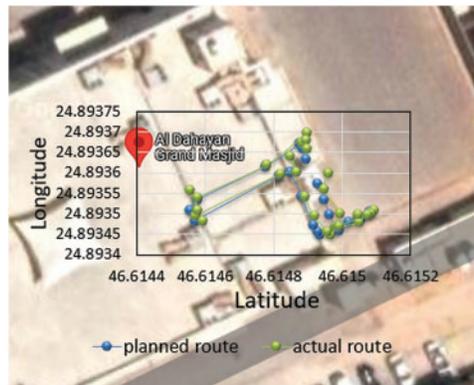


Figure 6: Navigation planned route vs. actual route

Table 8: Coordinates of outdoor navigation

Planned longitude	Planned latitude	Actual longitude	Actual latitude
24.89347	46.614989	24.893469	46.614989
24.89348	46.61502	24.89348	46.61505
24.89349	46.615048	24.893498	46.615048
24.89351	46.61508	24.893508	46.615092
24.8935	46.615083	24.8935	46.615083
24.89349	46.615056	24.893497	46.615076
24.89346	46.61499	24.89347	46.614991
24.89345	46.614936	24.893448	46.614966
24.89346	46.614928	24.893457	46.614958
24.89347	46.614915	24.893496	46.614918
24.89354	46.614881	24.893541	46.614891
24.8936	46.614847	24.893642	46.614847
24.8936	46.614847	24.893602	46.614867
24.8936	46.614818	24.893597	46.614819
24.89348	46.614572	24.893495	46.614572
24.89348	46.614572	24.893483	46.614592
24.89351	46.614557	24.893558	46.614557
24.89351	46.614557	24.893509	46.614578
24.89353	46.614578	24.89354	46.614578
24.89362	46.614776	24.893619	46.614779
24.89367	46.614877	24.893698	46.614897
24.89367	46.614877	24.893675	46.614882
24.89366	46.614884	24.89368	46.614894

5 Discussion

The results of this work indicate that the average response time for processing the command signal 0.5 s in order to avoid any accidents. The study also shows that the smart wheelchair program

can be used and applied without an internet connection. Moreover, the proposed program achieves a significant results in presence of external noise. As shown in Tab. 6, the accuracy, precision, recall, and F-Score for the implemented system have achieved an adequate results in comparing with the previous study in [26]. The experiment results revealed a statistically significant difference in the percentage values of the different categories, indicating a low risk of making incorrect predictions. For example, Tab. 7 shows that the difference between true and false predictions was about 187%. For the evaluation of the performance of indoor and outdoor navigation, the results indicated that the wheelchair was able to accurately maneuvering and RMSD was significantly low.

Although, this study enhances the system's suitability for a variety of users. However, wheelchairs require additional research in the static, motion, and moment of inertia domains. Additionally, the existing model of voice recognition omitted a speaker identification mechanism. By identifying a speaker, wheelchair users can only take particular directions from an authorized individual by increasing their safety. When comparing this study with other studies regarding efficacy, dependability, and cost, we believe that our design overcomes numerous complexities. For instance, in a recent study conducted by Abdulghani et al. [22], an adaptive neuro-fuzzy control was constructed and tested to track motorized wheelchairs using voice recognition. To achieve a high level of precision, the design must incorporate a wireless network in which the wheelchair is treated as a node. In another study, a wheelchair was driven using an eye and voice. In this study, the authors used a voice-controlled mode in conjunction with a web camera in order to make the system more congenial and reliable [33].

Despite this work has different merits; however some limitations need to be treated and updated in the following stages. For example, the system needs to be equipped with a variable controller and GPS to make it more efficient and meet the needs of users. In addition, the mechanical design of the wheelchair needs to be modified to change the torque of inertia. This change will alleviate the sudden jump of a wheelchair during the initial start or stop.

6 Conclusions

This research developed a voice-controlled wheelchair utilizing a low-cost and reliable technology. This technology uses a built-in voice recognition model combined with the CNN and LSTM models to train and classify five spoken commands. The design method used an Android smartphone (Flutter-based) app that connects with microcontrollers over an offline Wi-Fi hotspot. For the design and implementation of the experiment, ten native Arabic speakers produced a total of 2000 utterances of five words. The precision and usability of both indoor and outdoor navigation were tested using a range of disturbances. All voice commands have been given a normalized confusion matrix, precision, recall, and F-score. Voice recognition commands and wheelchair moves were demonstrated to be reliable in real-world testing. During indoor/outdoor maneuvering, it was also discovered that the RMSD estimated between the planned and real nodes was accurate. The ease of use, low cost, independence, and security are only a few of the benefits of the actual prototype. The device also has an emergency push-button as an additional safety element.

7 Future Work

The system can be enhanced with GPS technology, allowing users to design their own routes. The system also can be equipped with ultrasonic sensors for added safety, as it will operate and ignore user commands if the chair gets too close to an obstacle that could cause an accident. Additional research could be done to see if users prefer voice control interfaces over brain control interfaces. The voice

recognition model can be improved using the speaker identification algorithm to protect the disabled person's safety by accepting commands from only one user.

Acknowledgement: The author extend their appreciation to Deanship of Scientific Research, Majmaah University for supporting this work under Project Number (R-2022–77).

Funding Statement: This research was funded by the deputyship for Research and Innovation, Ministry of Education, Saudi Arabia, Grant Number IFP-2020–31.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] N. Vignier, J. F. Ravaud, M. Winance, F. X. Lepoutre and I. Ville, “Demographics of wheelchair users in France: Results of national community-based handicaps-incapacités-dépendance surveys,” *Journal of Rehabilitation Medicine*, vol. 40, no. 3, pp. 231–239, 2008.
- [2] Q. Zeng, C. L. Teo, B. Rebsamen and E. Burdet, “A collaborative wheelchair system,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 16, no. 2, pp. 161–170, 2008.
- [3] T. Carlson and Y. Demiris, “Collaborative control for a robotic wheelchair: Evaluation of performance, attention, and workload,” *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 42, no. 3, pp. 876–888, 2012.
- [4] J. Pineau, R. West, A. Atrash, J. Villemure and F. Routhier, “On the feasibility of using a standardized test for evaluating a speech-controlled smart wheelchair,” *International Journal of Intelligent Control and Systems*, vol. 16, no. 2, pp. 124–131, 2011.
- [5] A. Sharmila, A. Saini, S. Choudhary, T. Yuvaraja and S. G. Rahul, “Solar powered multi-controlled smart wheelchair for disabled: Development and features,” *Journal of Computational and Theoretical Nanoscience*, vol. 16, no. 11, pp. 4889–4900, 2019.
- [6] A. Hartman and V. K. Nandikolla, “Human-machine interface for a smart wheelchair,” *Journal of Robotics*, vol. 2019, pp. 11, 2019.
- [7] J. Tang, Y. Liu, D. Hu and Z. Zhou, “Towards BCI-actuated smart wheelchair system,” *Biomedical Engineering Online*, vol. 17, no. 1, pp. 1–22, 2018.
- [8] J. Leaman and H. M. La, “A comprehensive review of smart wheelchairs: Past, present, and future,” *IEEE Transactions on Human-Machine Systems*, vol. 47, no. 4, pp. 486–499, 2017.
- [9] G. Bourhis, O. Horn, O. Habert and A. Pruski, “An autonomous vehicle for people with motor disabilities,” *IEEE Robotics & Automation Magazine*, vol. 8, no. 1, pp. 20–28, 2001.
- [10] R. C. Simpson, “Smart wheelchairs: A literature review,” *Journal of Rehabilitation Research and Development*, vol. 42, no. 4, pp. 423, 2005.
- [11] S. Desai, S. S. Mantha and V. M. Phalle, “Advances in smart wheelchair technology,” in *Int. Conf. on Nascent Technologies in Engineering (ICNTE)*, Navi Mumbai, India, pp. 1–7, 2017.
- [12] Y. Rabhi, M. Mrabet and F. Fnaiech, “Intelligent control wheelchair using a new visual joystick,” *Journal of Healthcare Engineering*, vol. 2018, pp. 20, 2018.
- [13] S. Yathunathan, L. U. Chandrasena, A. Umakanthan, V. Vasuki and S. R. Munasinghe, “Controlling a wheelchair by use of EOG signal,” in *4th Int. Conf. on Information and Automation for Sustainability*, Sri Lanka, pp. 283–288, 2008.
- [14] B. Wiczorek, M. Kukla, D. Rybarczyk and Ł Warguła, “Evaluation of the biomechanical parameters of human-wheelchair systems during ramp climbing with the use of a manual wheelchair with anti-rollback devices,” *Applied Sciences*, vol. 10, no. 23, pp. 8757, 2020.

- [15] C. S. L. Tsui, P. Jia, J. Q. Gan, H. Hu and K. Yuan, "EMG-Based hands-free wheelchair control with EOG attention shift detection," in *2007 IEEE Int. Conf. on Robotics and Biomimetics (ROBIO)*, Sanya, China, pp. 1266–1271, 2007.
- [16] Y. Li, J. Pan, F. Wang and Z. Yu, "A hybrid BCI system combining p300 and SSVEP and its application to wheelchair control," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 11, pp. 3156–3166, 2013.
- [17] S. M. Hosni, H. A. Shedeed, M. S. Mabrouk and M. F. Tolba, "EEG-EOG based virtual keyboard: Toward hybrid brain computer interface," *Neuroinformatics*, vol. 17, no. 3, pp. 323–341, 2019.
- [18] S. D. Olesen, R. Das, M. D. Olsson, M. A. Khan and S. Puthusserypady, "Hybrid EEG-EOG-based BCI system for vehicle control," in *9th Int. Winter Conf. on Brain-Computer Interface (BCI)*, Gangwon, South Korea, pp. 1–6, 2021.
- [19] Z. T. Al-Qays, B. B. Zaidan, A. A. Zaidan and M. S. Suzani, "A review of disability EEG based wheelchair control system: Coherent taxonomy, open challenges and recommendations," *Computer Methods and Programs in Biomedicine*, vol. 1, no. 164, pp. 221–237, 2018.
- [20] N. Aktar, I. Jaharr and B. Lala, "Voice recognition based intelligent wheelchair and GPS tracking system," in *Int. Conf. on Electrical, Computer and Communication Engineering (ECCE)*, Cox's Bazar, Babgladesh, pp. 1–6, 2019.
- [21] Z. Raiyan, M. S. Nawaz, A. A. Adnan and M. H. Imam, "Design of an arduino based voice-controlled automated wheelchair," in *IEEE Region 10 Humanitarian Technology Conf. (R10-HTC)*, Dhaka, Babgladesh, pp. 267–270, 2017.
- [22] M. M. Abdulghani, K. M. Al-Aubidy, M. M. Ali and Q. J. Hamarsheh, "Wheelchair neuro fuzzy control and tracking system based on voice recognition," *Sensors*, vol. 20, no. 10, pp. 2872, 2020.
- [23] K. Sutikno Anam and A. Saleh, "Voice controlled wheelchair for disabled patients based on CNN and LSTM," in *4th Int. Conf. on Informatics and Computational Sciences (ICICoS)*, Semarang Indonesia, pp. 1–5, 2020.
- [24] O. Abdel-Hamid, A. R. Mohamed, H. Jiang, L. Deng, G. Penn *et al.*, "Convolutional neural networks for speech recognition," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 10, pp. 1533–1545, 2014.
- [25] M. S. I. Sharifuddin, S. Nordin and A. M. Ali, "Comparison of CNNs and SVM for voice control wheelchair," *IAES International Journal of Artificial Intelligence*, vol. 9, no. 3, pp. 387, 2020.
- [26] M. Bakouri, M. Alsehaimi, H. F. Ismail, K. Alshareef, A. Ganoun *et al.*, "Steering a robotic wheelchair based on voice recognition system using convolutional neural networks," *Electronics*, vol. 11, no. 1, pp. 168, 2022.
- [27] H. Alaeddine and M. Jihene, "Deep network in network," *Neural Computing and Applications*, vol. 33, pp. 1453–1465, 2021.
- [28] N. Kuzmin, K. Ignatiev and D. Grafov, "Experience of developing a mobile application using flutter," in *Information Science and Applications*, Singapore: Springer, pp. 571–575, 2020.
- [29] M. L. Napoli, *Beginning Flutter: A Hands on Guide to App Development*, New Jersey, USA: John Wiley & Sons, 2019.
- [30] M. S. B. A. Ghaffar, U. S. Khan, J. Iqbal, N. Rashid, A. Hamza *et al.*, "Improving classification performance of four class FNIRS-BCI using Mel Frequency Cepstral Coefficients (MFCC)," *Infrared Physics & Technology*, vol. 112, pp. 103589, 2021.
- [31] Y. J. Kim, J. P. Bae, J. W. Chung, D. K. Park, K. G. Kim *et al.*, "New polyp image classification technique using transfer learning of network-in-network structure in endoscopic images," *Scientific Reports*, vol. 11, no. 1, pp. 1–8, 2021.
- [32] G. V. Houdt, C. Mosquera and G. Nápoles, "A review on the long short-term memory model," *Artificial Intelligence Review*, vol. 53, no. 8, pp. 5929–5955, 2020.
- [33] S. Anwer, A. Waris, H. Sultan, S. I. Butt, M. H. Zafar *et al.*, "Eye and voice-controlled human machine interface system for wheelchairs using image gradient approach," *Sensors*, vol. 20, no. 19, pp. 5510, 2020.